# 3.0 VARIABILITY OF WET TEST METHODS

Chapter 3 describes the variability of effect concentration estimates (EC25, LC50, and NOEC) and endpoint measurements (survival, growth, and reproduction). For definitive studies of the variability of WET methods, readers should also refer to the TSD (USEPA 1991a, Part 1.3.3) and to WET methods manuals (USEPA 1993, 1994a, 1994b). EPA will complete and report on a new between-laboratory study of promulgated methods in 2000 or 2001.

## 3.1 Acquisition, Selection, and Quality Assurance of Data Presented in This Document

EPA solicited data for reference toxicant tests from laboratories that conduct WET tests and use reference toxicant testing as part of their quality control (QC) program. Reference toxicant testing is required, as specified in EPA toxicity test methods, to document laboratory performance over time for laboratories conducting self-monitoring tests. When laboratories are conducting effluent tests, at least one reference toxicant test must be conducted each month using the same toxicant, test concentrations, dilution water, and data analysis methods. These reference toxicant tests must be conducted using the same test conditions (type of dilution water, temperature, test protocol, and species) that are used for WET tests conducted by the laboratory.

Reference toxicant tests were used to characterize method variability because, in contrast to effluent samples, fixed concentrations of known toxicants are used. Only with this standardization is it possible to conclude that variability of the effect concentration estimates is derived from the sources discussed above, rather than from changes in the toxicant.

EPA received reference toxicant test data from several States, private laboratory sources, and the EPA Regions. Data sources used for these analyses include the EPA National Toxicant Reference Database (NTRD), the EPA Region 9 Toxicity Data Base, and laboratory bench sheets voluntarily submitted by independent sources. Although the data do not represent a random sample of laboratories or tests, they do represent a widespread sampling of typical laboratories and practices.

EPA required that reference toxicant tests included in its data base meet the following four criteria:

1. Test records documented the test method, organism, test date, laboratory, reference toxicant, and individual biological responses in the concentration series.

2. Data for each replicate were provided as required in the published method using the current test method.

3. The test used at least five toxicant concentrations and a control for the most commonly reported chronic toxicity test methods—(1) 1000.0, fathead minnow larval survival and growth; (2) 1002.0, *Ceriodaphnia* survival and reproduction; and (3) 1006.0, inland silverside survival and growth. For other chronic toxicity test methods, the test used at least four toxicant concentrations and a control because the methods permitted, in the recent past, the use of only four concentrations.

4. EPA personnel or an EPA contractor calculated the effect concentration, verified that all test acceptability criteria (TAC) had been met, and verified that the statistical flowchart had been followed correctly. Thus, all summary statistics and estimates were calculated from the replicate data and strictly followed the most current EPA test methods.

Details of data quality assurance and test acceptance are provided in a separate document, available at EPA's Office of Water docket, located in the Office of Science and Technology ["Whole Effluent Toxicity (WET) Data Test Acceptance and Quality Assurance Protocol"]. An attachment to that document provides a laboratory-by-laboratory listing of quality assurance flags, test dates, and toxicant concentrations, as well as summary statistics by laboratory for the NOEC, EC25, and LC50 estimates and test endpoints (survival, growth, reproduction, etc.). Laboratories are not identified by name.

The data set of reference toxicant tests includes information from 75 laboratories for 23 methods for tests conducted between 1988 and 1999. This document addresses, and provides specific guidance on, the variability of methods promulgated by EPA in 40 CFR Part 136 (Table 3-1). The data are also used to develop between-laboratory interim estimates of method variability for the promulgated methods (Appendix A). The Agency identifies these CVs as "interim;" EPA may revise some or all of these estimates based on between-laboratory studies to evaluate some of the promulgated test methods.

The next section presents summary statistics for the promulgated methods. Summary statistics for all methods in the data set appear in Appendix B. For methods represented by a few laboratories, summary statistics should not be considered representative of method performance. For example, EPA's Office of Water usually relies on acceptable data from at least six laboratories (USEPA 1996b) when it conducts a multi-laboratory study to quantify method performance. The data used here have not been obtained under conditions as rigorous as those applied to a between-laboratory study and for that reason, may overestimate variability, particularly for the extremes.

Coefficients of variation are used as descriptive statistics for NOECs in this document. Because NOECs can take on only values that correspond to concentrations tested, the distribution (and CV) of NOECs can be influenced by the selection of experimental concentrations, as well as additional factors (e.g., within-test variability) that affect both NOECs and point estimates. This makes CVs for NOECs more uncertain than the CVs for point estimates, and the direction of this uncertainty is not uniformly toward larger or smaller CVs. Despite these confounding issues, CVs are used herein as the best available means of expressing the variability of interest in this document and for general comparisons among methods. Readers should be cautioned, however, that small differences in CVs between NOECs and point estimates may be artifactual; large differences are more likely to reflect real differences in variability (a definition of what is "small" or "large" would require a detailed statistical analysis and would depend upon the experimental and statistical details surrounding each comparison). NOECs can only be a fixed number of discrete values; the mean, standard deviation, and CV cannot be interpreted and applied as they are for a continuous variable such as the EC25 or EC50. For instance, the typical reference toxicant test might result in only three observed NOEC values, most of them at one or two concentrations. The mean will fall between tested concentrations, as will the stated confidence intervals; thus, these do not actually represent expected outcomes, only approximations of the expected outcome.

As an alternative to CVs, ratios are used to quantify variability of EC25, EC50, and NOEC measurements in Appendix B. Ratios of measurements have been used previously to quantify and compare variability of NOEC and EC50 (Chapman et al. 1996b, Dhaliwal et al. 1997).

## 3.2 Variability of EC25, LC50, and NOEC

### *3.2.1 Within-Laboratory Variability of EC25, LC50, and NOEC*

This section characterizes the within-test and within-laboratory variability of effect concentration estimates. Tables 3-2 through 3-4 summarize variation across laboratories of the within-laboratory coefficients of variation (CVs), without respect to reference toxicant tested. Tables showing more extensive summaries appear in Appendix B (Tables B-1 through B-3).

**Table 3-1. Promulgated WET Methods Included in This Report**

| Test Method No. | Test Method | EPA Data Base | | |
|---|---|---|---|---|
| | | Toxicants | Tests | Labs |
| **Freshwater Methods for Chronic Toxicity**[a] | | | | |
| 1000.0 | *Pimephales promelas*, Fathead Minnow Larval Survival and Growth Test | Cd, Cr, Cu, KCl, NaCl, NaPCP, SDS | 205 | 19 |
| 1000.0 | *Pimephales promelas*, Fathead Minnow Embryo-Larval Survival and Teratogenicity Test | | 0 | 0 |
| 1002.0 | *Ceriodaphnia dubia*, Water Flea Survival and Reproduction Test | Cd, Cu, KCl, NaCl, NaPCP | 393 | 33 |
| 1003.0 | *Selenastrum capricornutum*,[b] Green Alga Growth Test | Cu, NaCl, Zn | 85 | 9 |
| **Marine & Estuarine Methods for Chronic Toxicity**[c] | | | | |
| 1004.0 | *Cyprinodon variegatus*, Sheepshead Minnow Larval Survival and Growth Test | Cd, KCl | 57 | 5 |
| 1005.0 | *Cyprinodon variegatus*, Sheepshead Minnow Embryo-larval Survival and Teratogenicity Test | | 0 | 0 |
| 1006.0 | *Menidia beryllina*, Inland Silverside Larval Survival and Growth Test | Cr, Cu, KCl, SDS | 193 | 16 |
| 1007.0 | *Americamysis (Mysidopsis) bahia*, Mysid Survival, Growth, and Fecundity Test | Cr, Cu, KCl | 130 | 10 |
| 1008.0 | *Arbacia punctulata*, Sea Urchin Fertilization Test | | 0 | 0 |
| 1009.0 | *Champia parvula*, Red Macroalga Reproduction Test | Cu, SDS | 23 | 2 |
| **Methods for Acute Toxicity** [d,e] | | | | |
| 2000.0 | Fathead Minnow Survival Test | Cd, Cu, KCl, NaCl, NaPCP | 217 | 21 |
| 2002.0 | *Ceriodaphnia dubia* Survival Test | Cd, Cu, KCl, NaCl, NaPCP | 241 | 23 |
| 2004.0 | Sheepshead Minnow Survival Test | SDS | 65 | 3 |
| 2006.0 | Inland Silverside Survival Test | Cd, KCl, SDS | 48 | 5 |
| 2007.0 | Mysid (*A. bahia*) Survival Test | Cd, Cu, SDS | 32 | 3 |
| 2011.0 | Mysid (*H. costata*) Survival Test | Cd, SDS | 14 | 2 |
| 2019.0 | Rainbow Trout Survival Test | Cu, Zn | 10 | 1 |
| 2021.0 | *Daphnia magna* Survival Test | Cd | 48 | 5 |
| 2022.0 | *Daphnia pulex* Survival Test | Cu, NaCl, SDS<br>Cd, Cu, NaCl, NaPCP | 57 | 6 |

[a] See publications EPA/600/4-89-001 (USEPA 1989) and EPA/600/4-91-002 (USEPA 1994b).

[b] The genus and species names for *Selenastrum capricornutum* have been changed to *Raphidocelis subcapitata*. In this document, however, *Selenastrum capricornutum* is used to avoid confusion.

[c] See publication EPA/600/4-91-003 (USEPA 1994a) and EPA/600/4-87/028 (USEPA 1988).

[d] See publications EPA/600/4-85/013 (USEPA 1985) and EPA/600/4-90/027F (USEPA 1993).

[e] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

Reference toxicant codes:

| | | | |
|---|---|---|---|
| Cd | cadmium | NaCl | sodium chloride |
| Cr | chromium | NaPCP | sodium pentachlorophenate |
| Cu | copper | SDS | sodium dodecyl sulfate |
| KCl | potassium chloride | Zn | zinc |

**Table 3-2. Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for EC25 (Chronic Tests)**

| Test Method[a] | Test Method No. | Endpoint[b] | No. of Labs | Percentiles of CV | | |
|---|---|---|---|---|---|---|
| | | | | 25th | 50th | 75th |
| Fathead Minnow Larval Survival & Growth | 1000.0 | G | 19 | 0.21 | 0.26 | 0.38 |
| Fathead Minnow Larval Survival & Growth | 1000.0 | S | 16 | 0.11 | 0.22 | 0.32 |
| *Ceriodaphnia* (Cd) Survival & Reproduction | 1002.0 | R | 33 | 0.17 | 0.27 | 0.45 |
| *Ceriodaphnia* (Cd) Survival & Reproduction | 1002.0 | S | 25 | 0.11 | 0.23 | 0.41 |
| Green Alga (*Selenastrum*) Growth | 1003.0 | G | 6 | 0.25 | 0.26 | 0.39 |
| Sheepshead Minnow Larval Survival & Growth | 1004.0 | G | 5 | 0.09 | 0.13 | 0.14 |
| Sheepshead Minnow Larval Survival & Growth | 1004.0 | S | 2 | 0.15 | 0.16 | 0.17 |
| Inland Silverside Larval Survival & Growth | 1006.0 | G | 16 | 0.18 | 0.27 | 0.43 |
| Inland Silverside Larval Survival & Growth | 1006.0 | S | 13 | 0.22 | 0.35 | 0.42 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | F | 4 | 0.30 | 0.38 | 0.41 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | G | 10 | 0.24 | 0.28 | 0.32 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | S | 7 | 0.17 | 0.21 | 0.28 |
| Red Macroalga (*Champia parvula*) Reproduction | 1009.0 | R | 2 | 0.58 | 0.58 | 0.59 |

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis* (*Mysidopsis*) *bahia*
[b] G = growth, S = survival, R = reproduction, F = fecundity

**Table 3-3. Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for LC50**

| Test Method[a] | Test Method No. | Endpoint[b] | No. of Labs | Percentiles of CV | | |
|---|---|---|---|---|---|---|
| | | | | 25th | 50th | 75th |
| **Freshwater Methods for Chronic Toxicity[c]** | | | | | | |
| Fathead Minnow Larval Survival & Growth | 1000.0 | S | 19 | 0.15 | 0.23 | 0.31 |
| *Ceriodaphnia* (Cd) Survival & Reproduction | 1002.0 | S | 33 | 0.10 | 0.16 | 0.29 |
| Sheepshead Minnow Larval Survival & Growth | 1004.0 | S | 5 | 0.07 | 0.08 | 0.12 |
| Inland Silverside Larval Survival & Growth | 1006.0 | S | 16 | 0.16 | 0.28 | 0.35 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | S | 10 | 0.16 | 0.26 | 0.27 |
| **Methods for Acute Toxicity[d,e]** | | | | | | |
| Fathead Minnow Larval Survival | 2000.0 | S | 21 | 0.10 | 0.16 | 0.19 |
| *Ceriodaphnia* (Cd) Survival | 2002.0 | S | 23 | 0.11 | 0.19 | 0.29 |
| Sheepshead Minnow Survival | 2004.0 | S | 5 | 0.12 | 0.14 | 0.21 |
| Inland Silverside Larval Survival | 2006.0 | S | 5 | 0.15 | 0.16 | 0.21 |
| Mysid (Ab) Survival | 2007.0 | S | 3 | 0.17 | 0.25 | 0.26 |
| Mysid (Hc) Survival | 2011.0 | S | 2 | 0.27 | 0.30 | 0.34 |
| Rainbow Trout Survival | 2019.0 | S | 1 | 0.23 | 0.23 | 0.23 |
| *Daphnia* (Dm) Survival | 2021.0 | S | 5 | 0.07 | 0.22 | 0.24 |
| *Daphnia* (Dp) Survival | 2022.0 | S | 6 | 0.19 | 0.21 | 0.27 |

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis* (*Mysidopsis*) *bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*
[b] S = survival
[c] See publications EPA/600/4-89-001 (USEPA 1989) and EPA/600/4-91-002 (USEPA 1994b).
[d] See publications EPA/600/4-85-013 (USEPA 1985 and EPA/600/4-90/027F (USEPA 1993).
[e] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

**Table 3-4. Quartiles (25th and 75th) and Median (50th) of the Within-Laboratory Values of CV for NOEC**

| Test Method[a] | Test Method No. | Endpoint[b] | No. of Labs | Percentiles of CV | | |
|---|---|---|---|---|---|---|
| | | | | 25th | 50th | 75th |
| **Freshwater Methods for Chronic Toxicity[c]** | | | | | | |
| Fathead Minnow Larval Survival & Growth | 1000.0 | G | 19 | 0.22 | 0.37 | 0.53 |
| Fathead Minnow Larval Survival & Growth | 1000.0 | S | 19 | 0.26 | 0.39 | 0.48 |
| *Ceriodaphnia* (Cd) Survival & Reproduction | 1002.0 | R | 33 | 0.25 | 0.33 | 0.49 |
| *Ceriodaphnia* (Cd) Survival & Reproduction | 1002.0 | S | 33 | 0.21 | 0.30 | 0.43 |
| Green Alga (*Selenastrum*) Growth | 1003.0 | G | 9 | 0.40 | 0.46 | 0.56 |
| **Marine & Estuarine Methods for Chronic Toxicity[d]** | | | | | | |
| Sheepshead Minnow Larval Survival & Growth | 1004.0 | G | 5 | 0.34 | 0.40 | 0.44 |
| Sheepshead Minnow Larval Survival & Growth | 1004.0 | S | 5 | 0.14 | 0.18 | 0.24 |
| Inland Silverside Larval Survival & Growth | 1006.0 | G | 16 | 0.31 | 0.46 | 0.57 |
| Inland Silverside Larval Survival & Growth | 1006.0 | S | 16 | 0.30 | 0.42 | 0.55 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | F | 4 | 0.17 | 0.36 | 0.40 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | G | 10 | 0.35 | 0.39 | 0.43 |
| Mysid (Ab) Survival, Growth, & Fecundity | 1007.0 | S | 10 | 0.28 | 0.33 | 0.38 |
| Red Macroalga (*Champia parvula*) Reprod. | 1009.0 | R | 2 | 0.85 | 1.00 | 1.16 |
| **Methods for Acute Toxicity[e,f]** | | | | | | |
| Fathead Minnow Larval Survival | 2000.0 | S | 21 | 0.18 | 0.22 | 0.34 |
| *Ceriodaphnia* (Cd) Survival | 2002.0 | S | 23 | 0.18 | 0.35 | 0.41 |
| Sheepshead Minnow Survival | 2004.0 | S | 3 | 0 | 0.31 | 0.33 |
| Inland Silverside Larval Survival | 2006.0 | S | 5 | 0 | 0.33 | 0.35 |
| Mysid (Ab) Survival | 2007.0 | S | 3 | 0.29 | 0.38 | 0.43 |
| Mysid (Hc) Survival | 2011.0 | S | 2 | 0.21 | 0.26 | 0.31 |
| Rainbow Trout Survival | 2019.0 | S | 1 | 0.35 | 0.35 | 0.35 |
| *Daphnia magna* (Dm) Survival | 2021.0 | S | 5 | 0.09 | 0.36 | 0.47 |
| *Daphnia pulex* (Dp) Survival | 2022.0 | S | 6 | 0.21 | 0.38 | 0.61 |

[a] Cd = *Ceriodaphnia dubia*, Ab = *Americamysis* (*Mysidopsis*) *bahia*, Hc = *Holmesimysis costata*, Dm = *Daphnia magna*, Dp = *Daphnia pulex*

[b] G = growth, S = survival, R = reproduction, F = fecundity

[c] See publications EPA/600/4-89-001 (USEPA 1989) and EPA/600/4/4-91-002 (USEPA 1994b).

[d] See publication EPA/600/4-91-003 (USEPA 1994a) and EPA/600/4-87/028 (USEPA 1988).

[e] See publications EPA/600/4-85/013 (USEPA 1985) and EPA/600/4-90/027F (USEPA 1993).

[f] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

Effect concentrations having a p-percent effect are symbolized as ECp and may be calculated for sublethal and lethal (survival) endpoints (USEPA 1993,1994a,1994b). Effect concentrations commonly estimated for WET methods are LC50, EC50, IC25, and EC25. The symbol ECp is more general and may be used to represent an LCp, ECp, or ICp endpoint. To simplify presentation of results in this document, the term EC25 is used to represent the concentration at which a 25-percent effect has occurred for either lethal

or sublethal endpoints. The term LC50 is used to represent the concentration at which a 50-percent effect has occurred for lethal endpoints. The EC25 for survival is not routinely used in generating self-monitoring data and is presented here for comparison to the EC25 for sublethal endpoints (i.e., IC25). Estimates of EC25, LC50, and NOEC were calculated for this document as required in the EPA test methods (USEPA 1993, 1994a, 1994b). A CV is reported for NOEC measurements in this document. See Appendix A for further details.

The results in Tables 3-2 through 3-4 were obtained as follows, using as an example the EC25 of the growth endpoint in Method 1000.0 (fathead minnow larval chronic test) on the first row of Table 3-2. The CV of the EC25 estimates was calculated for each laboratory. This calculation resulted in 19 CVs (one per laboratory with each laboratory tested using one toxicant). The sample percentiles were calculated for this set of 19 CVs. In Table 3-2, the column headed "$50^{th}$" shows the $50^{th}$ percentile (median value) of CV found across these 19 laboratories; the $50^{th}$ percentile value is 0.26. In the column headed "$75^{th}$," the $75^{th}$ percentile CV is reported as 0.38. When a method is represented by fewer than four laboratories, the minimum and maximum CVs are shown in the columns headed "$25^{th}$" and "$75^{th}$," respectively. Note that these CVs represent within-laboratory variability, and that Tables 3-2 through 3-4 show the quartiles and median of the within-laboratory CVs. These tables thus report the typical range of within-laboratory test method variation.

Variation across laboratories in the CV for effect concentration estimates (Tables 3-2 through 3-4) may be summarized as follows, ignoring methods represented by only one or two laboratories. [Refer to the column headed "$75^{th}$" (the $75^{th}$ percentile).]

For the EC25 of the growth and reproduction endpoints in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.14 to 0.45 depending on the method (Table 3-2). For the two most commonly used methods (1000.0, fathead minnow larval chronic test; and 1002.0, *Ceriodaphnia* chronic test), 75 percent of the laboratories have CVs no more than 0.38 and 0.45, respectively.

For the LC50 of the survival endpoint in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.12 to 0.35, depending on the method. For the two most commonly used methods (1000.0 and 1002.0), 75 percent of laboratories have CVs no more than 0.31 and 0.29, respectively (Table 3-3). For the LC50 in acute toxicity tests, 75 percent of laboratories have a CV no more than 0.19 to 0.29, depending on the method. For the two most commonly used methods (2000.0 and 2002.0), 75 percent of laboratories have CVs no more than 0.19 and 0.29, respectively.

For the NOEC of growth or reproduction endpoints in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.43 to 0.57, depending on the method. For the two most commonly used methods (1000.0 and 1002.0), 75 percent of laboratories have CVs no more than 0.53 and 0.49, respectively (Table 3-4). For the NOEC of survival in chronic toxicity tests, 75 percent of laboratories have a CV no more than 0.24 to 0.55, depending on the method. For the two most commonly used methods (1000.0 and 1002.0), 75 percent of laboratories have CVs no more than 0.48 and 0.43, respectively. For the NOEC of survival in acute toxicity tests, 75 percent of laboratories have a CV no more than 0.34 to 0.61, depending on the method. For the two most commonly used acute methods (2000.0 and 2002.0), 75 percent of laboratories have CVs no more than 0.34 and 0.41, respectively.

Appendix B discusses the range of toxicant concentrations reported as the NOEC. For chronic toxicity tests, most laboratories report the NOEC to within two to three concentration intervals, and half the laboratories report most NOECs within one to two concentration intervals for reference toxicants. For acute toxicity tests, most laboratories report NOECs at one or two concentrations. This outcome agrees with EPA's expected performance for these methods. The normal variation of the effect concentration estimate in reference toxicant tests has been reported for some EPA WET methods (USEPA 1994a, 1994b) to be plus or minus one dilution concentration for the NOEC and less for LC50.

### 3.2.2 Between-Laboratory Variability of EC25, LC50, and NOEC

The data set compiled for this document provided reasonable estimates of between-laboratory variability for only a few methods. For many methods and toxicants, there were too few laboratories in the data base. Additional summaries of between-laboratory variability of WET methods are included in the TSD (USEPA 1991a, Part 1.3.3) and the WET methods manuals (USEPA 1994a, 1994b). EPA also intends to provide new data in a forthcoming EPA between-laboratory study of promulgated methods.

Using the data set, credible estimates of between-laboratory variability could be made for a few toxicants and methods having data for six or more laboratories (Table 3-5). The statistical methods are described in Appendix B. Table 3-5 shows values of the square root of within-laboratory and between-laboratory variance components (i.e., standard deviations, σ). The standard deviations and mean are expressed in units of toxicant concentration (e.g., g/L or mg/L). Between-laboratory $\sigma_b$ estimates the standard deviation for laboratory means of EC25, LC50, and NOEC. The "Mean" column in Table 3-5 shows the mean of the laboratory means, not the mean for all tests. Because the number of tests differed among laboratories, these two means are different. These data suggest that between-laboratory variability ($\sigma_b$) is comparable to within-laboratory variability ($\sigma_w$) for the methods listed in the table.

In Table 3-5, the ratio of $\sigma_b$ to the mean is an estimate of the relative variability ($CV_b$) of laboratory means around their combined mean. The ratio of $\sigma_w$ to the mean may approach the value of the average within-laboratory CV when the sample of laboratories is large, but to characterize within-laboratory CVs, readers should use Tables 3-2 through 3-4.

**Table 3-5. Estimates of Within-Laboratory and Between-Laboratory Components of Variability[a]**

| Test Method[b] | Test EC Estimate | Toxicant | End-Point[c] | Tests | Labs | Within-lab $\sigma_w$ | Between-lab $\sigma_b$ | Mean | $CV_w$ | $CV_b$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1000.0 | EC25 | NaCl | G | 73 | 6 | 0.67 | 0.44 | 2.63 | 0.25 | 0.17 |
| 1000.0 | LC50 | NaCl | S | 73 | 6 | 1.14 | 0.45 | 4.15 | 0.27 | 0.11 |
| 1000.0 | NOEC | N Cl | G | 73 | 6 | 0.72 | 0.35 | 2.18 | 0.33 | 0.16 |
| 1000.0 | NOEC | NaCl | S | 73 | 6 | 0.96 | 0.51 | 2.43 | 0.40 | 0.21 |
| 1002.0 | EC25 | NaCl | R | 292 | 23 | 0.29 | 0.27 | 0.92 | 0.32 | 0.29 |
| 1002.0 | LC50 | NaCl | S | 285 | 23 | 0.48 | 0.24 | 1.78 | 0.27 | 0.13 |
| 1002.0 | NOEC | NaCl | G | 292 | 23 | 0.28 | 0.18 | 0.74 | 0.38 | 0.24 |
| 1002.0 | NOEC | NaCl | S | 292 | 23 | 0.47 | 0.26 | 1.42 | 0.33 | 0.18 |
| 1006.0 | EC25 | Cu | G | 130 | 9 | 45.1 | 52.4 | 97.4 | 0.46 | 0.54 |
| 1006.0 | LC50 | Cu | S | 130 | 9 | 48.4 | 70.7 | 127.0 | 0.38 | 0.56 |
| 1006.0 | NOEC | Cu | G | 130 | 9 | 51.8 | 44.4 | 80.1 | 0.65 | 0.55 |
| 1006.0 | NOEC | Cu | S | 130 | 9 | 34.2 | 39.5 | 65.4 | 0.52 | 0.60 |
| 2000.0 | LC50 | NaCl | S | 154 | 14 | 1.05 | 1.24 | 7.46 | 0.14 | 0.17 |
| 2002.0 | LC50 | NaCl | S | 167 | 15 | 0.36 | 0.38 | 1.97 | 0.18 | 0.19 |

[a] $\sigma_w$ = within-laboratory standard deviation, $\sigma_b$ = between-laboratory standard deviation
$CV_w$ = within-laboratory coefficient of variation, $CV_b$ = between-laboratory coefficient of variation

[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

[c] G = growth, S = survival, R = reproduction

## 3.3 Variability of Endpoint Measurements

This section characterizes the within-laboratory precision of endpoint measurements (e.g., growth, reproduction, and survival). Endpoint variability in methods for chronic toxicity is characterized here using sublethal endpoints. The sublethal endpoint was designed to be more sensitive than the survival endpoint, and it incorporates the effect of mortality (i.e., it incorporates biomass). For example, for the chronic survival and growth fathead minnow larval test, the total dry weight at each replicate is divided by the original number of larvae, rather than the surviving number of larvae.

EPA reports measures of test precision based on the control CV [(control standard deviation)/(control mean)] and the "Percent MSD" [100×MSD/(control mean)], symbolized as PMSD. Recall that MSD, the "minimum significant difference," is calculated as $[d \sqrt{EMS} \sqrt{(2/r)}]$, where "d" is the critical value of Dunnett's statistic when comparing "k" treatments to a control, EMS is the error mean square from the analysis of variance of the endpoint responses, and "r" is the number of replicates at each concentration (USEPA 1993, 1994a, 1994b). These measures of test precision quantify within-test variability, or the sensitivity of each test to toxic effects on the biological endpoint.

Measures of variability relative to the control mean are used for two reasons. First, a laboratory having consistently large mean endpoint values for the control will also tend to have larger values of MSD and control standard deviation. Second, PMSD is readily interpreted as the minimum percent difference between control and treatment that can be declared statistically significant in a WET test. A significant effect occurs when (control mean - treatment mean) exceeds the MSD. Dividing by the control mean and multiplying by 100 states this relationship in terms of the percent difference between control and treatment.

To characterize the distribution of values of PMSD, values from all laboratories and toxicants for a given method and endpoint were combined, and sample percentiles reported. Percentiles are also reported for the CV of the control, which also indicates variability among replicates under non-toxic conditions and may be a useful indicator of uniformity of the test organisms. The sample percentiles are reported in more detail in Appendix B; the $10^{th}$ and $90^{th}$ percentiles are shown in Table 3-6. Method 1009.0 (red macroalga) is omitted from Table 3-6 because it would be inadvisable to characterize method variability using only 23 tests from only two laboratories.

The $90^{th}$ percentile may be used as an upper PMSD bound (i.e., a limit on the insensitivity of a test). The $10^{th}$ percentile may be used as a lower PMSD bound for declaring a significant difference or a lower limit to test sensitivity. The $90^{th}$ percentile has been used in other WET programs (Chapter 5). The $95^{th}$ percentile is used as a practical upper limit for the variability of analytical results in well-controlled between-laboratory studies that use a standard protocol and specific quality assurance procedures (ASTM 1992, 1998; USEPA 1993, 1996a, 1996b). The tests summarized here have not been subjected to the rigorous standardization and quality assurance of collaborative studies, and the data have not been screened for outliers as specified by ASTM Practices D2777 and E691 (ATSM 1992, 1998). These considerations justify using the sample $90^{th}$ percentile to set an upper bound. A lower bound is necessary to avoid creating a disincentive for improving test precision and to objectively specify a limit to the test sensitivity achieved in practice. If no more than ten percent of tests are more precise than this lower bound, then in practice, the analytical method rarely detects toxic effects of this small magnitude.

When comparing values in Table 3-6 to a test result, it is important that the test's MSD be calculated according to procedures described in the EPA method manuals (USEPA 1993, 1994a, 1994b) for Dunnett's test for multiple comparisons with a control (see Section 6.4.1). An analysis of variance (ANOVA) is conducted using several treatments, including the control. EPA methods require excluding from the ANOVA those concentrations for which no organisms survived in any replicate. For a sublethal endpoint, concentrations are excluded from the analysis if they exceed the NOEC for survival. The MSD is calculated

using the square root of the error mean square (rEMS) from the ANOVA, and using Dunnett's critical value (which depends on the number of replicates and concentrations used in the ANOVA).

**Table 3-6. Range of Relative Variability for Endpoints of Promulgated WET Methods, Defined by the 10th and 90th Percentiles from the Data Set of Reference Toxicant Tests[a]**

| Test Method[b] | Endpoint[c] | No. of Labs | No. of Tests | PMSD | | Control CV[d] | |
|---|---|---|---|---|---|---|---|
| | | | | 10th | 90th | 10th | 90th |
| 1000.0 Fathead Minnow | G | 19 | 205 | 9.4 | 35 | 0.035 | 0.20 |
| 1002.0 *Ceriodaphnia dubia* | R | 33 | 393 | 11 | 37 | 0.089 | 0.42 |
| 1003.0 Green Alga | G | 9 | 85 | 9.3 | 23 | 0.034 | 0.17 |
| 1004.0 Sheepshead Minnow | G | 5 | 57 | 6.3 | 23 | 0.034 | 0.13 |
| 1006.0 Inland Silverside | G | 18 | 193 | 12 | 35 | 0.044 | 0.18 |
| 1007.0 Mysid | G | 10 | 130 | 12 | 32 | 0.088 | 0.28 |
| 2000.0 Fathead Minnow | S | 20 | 217 | 4.2 | 30 | 0 | 0.074 |
| 2002.0 *Ceriodaphnia* | S | 23 | 241 | 5.0 | 21 | 0 | 0.11 |
| 2004.0 Sheepshead Minnow | S | 5 | 65 | 0[e] | 55 | 0 | 0 |
| 2006.0 Inland Silverside | S | 5 | 48 | 7.0 | 41 | 0 | 0.079 |
| 2007.0 Mysid (*A. bahia*) | S | 3 | 32 | 5.1 | 26 | 0 | 0.081 |
| 2011.0 Mysid (*H. costata*) | S | 2 | 14 | 18 | 47 | 0 | 0.074 |
| 2021.0 Daphnia (*D. magna*) | S | 5 | 48 | 5.3 | 23 | 0 | 0.11 |
| 2022.0 Daphnia (*D. pulex*) | S | 6 | 57 | 5.8 | 23 | 0 | 0.11 |

[a] The precision of the data warrants only three significant figures. When determining agreement with these values, one may round off values to two significant figures (e.g., values >3.45000... and ≤3.5000... are rounded to 3.5). Method 1009.0 (red macroalga) is not reported because it is inadvisable to characterize method variability using only 23 tests from just two laboratories.

[b] EPA did not assign method numbers for acute methods in EPA/600/4-90/027F. The numbers assigned here were created for use in this document and in related materials and data bases.

[c] G = growth, R = reproduction, S = survival

[d] CVs were calculated using untransformed control means for each test.

[e] An MSD of zero will not occur when the EPA flow chart for statistical analysis is followed. In this report, MSD was calculated for every test, including those for which the flow chart would require a nonparametric hypothesis test. EPA recommends using the value 4.2 (the 10th percentile shown for the fathead minnow acute test) in place of zero as the 10th percentile PMSD (lower PMSD bound) for the sheepshead minnow acute test.

The MSD was calculated for all test results reported here, including those for which non-normality and heterogeneity of variance were indicated. Thus, this document presents MSD as an approximate index of test sensitivity. Estimates of power are also approximate. The MSD generally will be related to test sensitivity, even when the assumptions for ANOVA and Dunnett's test are not strictly satisfied.

Table 3-7 shows the number of laboratories in the WET variability data set having tests exceeding the upper PMSD bound reported in Table 3-6. One-half to two-thirds of the laboratories never or infrequently exceeded the bound, and roughly one in five exceeded it in at least 20 percent of their tests. By definition of the 90th percentile, about 10 percent of all the tests exceeded the bound.

**Table 3-7. Number of Laboratories Having a Given Percent of Tests Exceeding the PMSD Upper Bound for the Sublethal Endpoint**

| Test Method | No. Labs | Endpoints[a] | Number of Labs with Various Percentages of Tests Exceeding the PMSD Upper Bound | | | | |
|---|---|---|---|---|---|---|---|
| | | | 0% | 0%-10% | 10%-20% | 20%-50% | 50%-100% |
| 1000.0 Fathead Minnow | 19 | G | 8 | 2 | 7 | 2 | 0 |
| 1002.0 *Ceriodaphnia dubia* | 33 | R | 15 | 7 | 5 | 6 | 0 |
| 1003.0 Green Alga | 9 | G | 6 | 1 | 0 | 2 | 0 |
| 1004.0 Sheepshead Minnow | 5 | G | 3 | 1 | 0 | 1 | 0 |
| 1006.0 Inland Silverside | 16 | G | 6 | 5 | 1 | 4 | 0 |
| 1007.0 Mysid (growth) | 10 | G | 5 | 2 | 0 | 3 | 0 |

[a] G = growth, R = reproduction

## 3.4 Conclusions about Variability of WET Methods

### *3.4.1 Variability of EC25, LC50, NOEC*

For EC25, the quartiles of the within-laboratory CVs ranged across the promulgated methods from 0.09 to 0.45, and the median CV ranged from 0.13 to 0.38. For LC50, the quartiles of the within-laboratory CVs ranged from 0.07 to 0.35, and the median CV ranged from 0.08 to 0.28. For NOEC, the quartiles of the within-laboratory CVs ranged from 0 to 0.61, and the median CV ranged from 0.18 to 0.46. This summary applies to those methods represented by at least 20 tests and three laboratories.

EPA concludes from Tables 3-2 through 3-4 that point estimates are substantially less variable than the NOEC for the same method and endpoint, and that the LC50 for an acute toxicity test usually is less variable than the LC50 for a chronic toxicity test. The estimated NOEC is more variable than ECp *using current experimental designs* because NOEC can take only those values equal to the concentrations tested, while ECp interpolates between tested concentrations (there may be other, more technical reasons as well). In principle, NOEC could be estimated more accurately and precisely by changing the experimental design to use more concentrations at narrower dilution ratios and by using more replicates. The greater variability of the NOEC underscores the desirability of using point estimates to characterize effluent toxicity.

Tables 3-2 through 3-4 may be used as benchmarks for variability, allowing comparison of one laboratory's CV for reference toxicant testing with CVs reported by experienced laboratories reporting tests that passed the TAC. However, CVs for methods represented by too few laboratories in the table may be atypical.

The CVs in Tables 3-2 through 3-4 may be used as an adjunct to the control chart. If the CV for reference toxicant tests is above the 75$^{th}$ percentile in Tables 3-2 through 3-4, variability likely can be reduced, even if the individual EC25 or LC50 values fall within the control limits. If a control chart is constructed using an unreasonably large standard deviation, the control limits will be unreasonable. If a high CV is not fully explained by an unusually small mean, the standard deviation of EC25 or LC50 should be reduced to bring the CV within the normal range. If the CV exceeds the 90$^{th}$ percentile (Appendix B), there is no question that variability is unacceptably large. Detailed guidance is provided in Chapter 5 (Section 5.3.1.1).

Tables 3-2 through 3-4 indicate the magnitude of the analytical variability that becomes part of the variability of effluent test results under certain conditions. This occurs when effluent test results (NOECs, LC50s, or EC25s) fall between the lowest and highest concentrations tested. Under other conditions, these

CVs may not accurately represent analytical variability. If tests give results consistently near or at the lowest or highest concentrations tested, or if the tests often produce "less than" or "greater than" results, Tables 3-2 through 3-4 will not accurately characterize the analytical CV for such tests. To measure the analytical CV under such conditions, reference toxicant tests would have to be designed to have the effect concentration at or near the lowest or highest concentration. The CV and standard deviation measured under such conditions are unknown, but are likely to differ from those for standard reference toxicant tests.

The data set did not contain information supporting an analysis of the causes of between-laboratory variability. Possible causes may include laboratory differences in concentration series, incorrect or ambiguous calculation or reporting of concentrations (e.g., concentration of the metal ion versus the salt), laboratory differences in dilution water (e.g., water hardness or pH), laboratory differences in foods and feeding regimes, and laboratory differences in cultures (genotypic and phenotypic differences in sensitivity to various toxicants).

The lack of a standard or common reference toxicant creates a problem for permittees and regulatory authorities attempting to evaluate and compare laboratories. Real or apparent differences occur between laboratories in the mean values of EC25, LC50, and NOEC. Some of this difference is random and reflects only the within-laboratory variance; some may be systematic. Systematic, between-laboratory differences can be inferred reliably only when laboratories use the same test method, use the same reference toxicants and dilution series, use similar dilution waters, and report a sufficient number of tests.

### *3.4.2 Variability of Endpoint Measurements*

EPA has selected the PMSD to characterize endpoint variability for WET test methods because it integrates variability from several concentrations (always including the control), and it represents the MSD used in the WET hypothesis test. The control CV, by itself, does not fully represent the variability affecting a WET hypothesis test or point estimate. The PMSD also represents the variability affecting point estimates because it is calculated using the EMS for the endpoint measurement. (However, the standard error of a point estimate of an effect concentration may be a complicated function of the EMS.)

PMSD for sublethal endpoints ranged from 6 to 37 across the promulgated chronic methods. For the fathead minnow chronic method, PMSD ranged from 9 to 35; for the *Ceriodaphnia* chronic method, PMSD ranged from 11 to 37. Thus, most chronic tests were able to distinguish a reduction of 37 percent or smaller in the endpoint. Further analysis in Chapter 5 shows that most tests were unable to distinguish consistently a 25-percent reduction. For the survival endpoint of promulgated acute methods, PMSD ranged from 0 to 55. For the two most commonly used acute methods (fathead minnow and *Ceriodaphnia*), PMSD ranged from 4 to 30 and from 5 to 21, respectively. Thus, PMSD varied markedly for some acute methods and not for others.

As shown by the size of PMSD, test sensitivity to detect substantial toxic effects is occasionally insufficient at some laboratories and routinely insufficient at a few laboratories. Inadequate test sensitivity is not always signaled by control charts of EC25, LC50, and NOEC. Laboratories should consider maintaining control charts for MSD or PMSD, and should report MSD and the control mean with all WET tests.

Some portion of MSDs in the WET variability data set could be considered exceptionally large, if not outliers. This observation underscores the importance of a careful review for each WET test, including an examination of means and standard deviations for endpoint responses at each concentration; the plotting of replicate data (not just concentration means); and, when necessary, a search for possible causes of excessive variability. The tables and plots in the promulgated methods (USEPA 1994a, 1994b) provide good examples.

***This page intentionally left blank.***